JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 1, 113—120 (1960)

Ergodic Theorem for General Functions*

JOHN G. KEMENY

Department of Mathematics, Dartmouth College, Hanover, New Hampshire

The purpose of this paper is to prove a generalization of the classical ergodic theorem, which is applicable to functions with non-numerical values. We will proceed by giving a suggestive reformulation of the classical result, passing to the abstract formulation, and then proving the generalized theorem.

Given a measure space $(\boldsymbol{X},\mathbf{S},\mu)$, a real-valued integrable function f, and a measure preserving transformation T, the ergodic theorem asserts that the averages

$$a_n(x) = (1/n)[f(x) + f(Tx) + \ldots + f(T^{n-1}x) \tag{1}$$

converge a.e. (When we say "almost everywhere" or "a.e.", we will always mean that the event occurs for all $x \in \boldsymbol{X}$ except for a set of measure 0.) We will be concerned only with totally finite measures and will choose $\mu(\boldsymbol{X}) = 1$.

The classical theorem asserts (for the totally finite case) the existence of a function $a(x)$ such that

(I) $a_n(x) \to a(x)$ a.e.

(II) $a(Tx) = a(x)$ a.e.

(III) $\int a\, d\mu = \int f\, d\mu$.

If T is ergodic (metrically transitive), then we also know that

(IV) $a(x) = \int f\, d\mu$ a.e.

---

* The preparation of this paper was aided by the National Science Foundation through a grant to the Dartmouth Mathematics Project. The author also wishes to express his thanks to J. L. Snell and R. E. Williamson for their helpful suggestions.

Let us rewrite (1) as an integral, by introducing the measures $\mu_{n,x}$ which put weight $1/n$ at each of the points $T^i x$, $i = 0,1,\ldots, n-1$. Then

$$a_n(x) = \int f\, d\mu_{n,x}. \tag{1'}$$

Our only numerical operation in (1′), (I)—(IV) is integration with respect to a measure μ, with $\mu(\boldsymbol{X}) = 1$. This is a mean, and we will make use of the fact that the mean minimizes the variance. Hence we introduce the variance relative to a real number t:

$$V_\mu(f,t) = \int (f(x) - t)^2\, d\mu(x). \tag{2}$$

The assumption that these variances are finite is stronger than the assumption that f is integrable. But this, stronger assumption can be abstracted, integrability cannot.

DEFINITION 1. The mean $m_\mu(f)$ is the number for which $V_\mu(f,\cdot)$ takes on its minimum value.

In the classical case $m_\mu(f) = \int f\, d\mu$. We may now reformulate (I)—(IV) as follows: There exists a function a such that

(I′) $m_{\mu_{n,x}}(f) \to a(x)$ a.e.

(II′) $a(Tx) = a(x)$ a.e.

(III′) $m_\mu(a) = m_\mu(f)$.

If T is ergodic

(IV′) $a(x) = m_\mu(f)$ a.e.

(I′)—(IV′) make sense for non-numerical functions f, as long as we can interpret means. This in turn (see Def. 1) requires the introduction of a real-valued function $V_\mu(f,\cdot)$ for every μ and f.

We are now ready for the abstract formulation. We start with the measure space $(\boldsymbol{X},\mathbf{S},\mu)$, and the measure preserving transformation T. But f has a range $\boldsymbol{R}$ that may be an arbitrary set. We also need a space $\boldsymbol{M}$ from which means are selected. We certainly don't want to require that $\boldsymbol{R} = \boldsymbol{M}$, since, for example, integer valued functions may have non-integer means. Hence $\boldsymbol{R} \subset \boldsymbol{M}$ is suggested, and this will be the case in most applications. But we won't even make this restriction. The set $\boldsymbol{R}$ is arbitrary, and $\boldsymbol{M}$ is any topological space.

Next we must assign a variance function $V_\nu(g,\cdot)$ to each measure ν on $\boldsymbol{X}$ with total measure 1, and each function g on $\boldsymbol{X}$ with range in $\boldsymbol{R}$.

We will do this in analogy to the classical case where we will think of $\boldsymbol{R}$ and $\boldsymbol{M}$ as *bounded* sets of real numbers. Then $V_\nu(g, \cdot)$ is a continuous function on $\boldsymbol{M}$. Actually we will need only that

(a) $V_\nu(g, \cdot)$ is a bounded, non-negative real-valued function on $\boldsymbol{M}$, that takes on a minimum in every closed subset of $\boldsymbol{M}$.

For example, if $\boldsymbol{M}$ is compact, the lower semi-continuity of the functions will suffice. We must next connect the measure with its variance. In a certain sense V_ν depends continuously on ν. This will be made precise as follows: Assume that

(b) $\mathbf{A} = \{S_i\}$ is a denumerable algebra of measurable sets, such that $\mathbf{S}$ is the σ-algebra generated by $\mathbf{A}$.

Then we introduce a topology on our measures.

Definition 2. $\nu_n \to \nu$ if for all $S_i \in \mathbf{A}$, $\nu_n(S_i) \to \nu(S_i)$.

The topology to be used on the bounded functions on $\boldsymbol{M}$ will be given by a norm:

Definition 3. If h is bounded on $\boldsymbol{M}$,

$$|h| = \sup_{t \in \boldsymbol{M}} |h(t)|.$$

Then we assume

(c) For fixed g, the mapping $\nu \to V_\nu(g, \cdot)$ is continuous.

We should check that this holds in a typical classical case. Let our measure space be $[0,1]$, with the Borel sets as measurable sets. Let $\mathbf{A}$ be the algebra generated by intervals with rational end-points. Then (b) is satisfied. If f is continuous, then (2) will satisfy (a). And if $\nu_n(S_i) \to \nu(S_i)$ for all rational intervals S_i, then $\int (f(x) - t)^2 \, d\nu_n(x) \to \int (f(x) - t)^2 \, d\nu$, uniformly in t. Hence (c) holds.

While this argument establishes the reasonableness of Def. 2, we must still show that the simpler requirement — that $\nu_n(S_i) \to \nu(S_i)$ for *all* measurable sets S_i — is not acceptable. We will return to the example of the interval $[0,1]$. We will choose Lebesgue measure, and an ergodic transformation T. It will follow from Lemma 3 that $\mu_{n,x}(S_i) \to \mu(S_i)$, for a fixed denumerable class of measurable sets, a.e. But *no* sequence $\mu_{n,x}$ converges to the right answer for all S_i. Specifically, let $S_x = \{T^i x\}$, $i = 0,1, \ldots$. This set is denumerable, hence its Lebesgue measure is 0. But $\mu_{n,x}(S_x) = 1$.

We would like to apply Def. 1, to get a mean point $m_\mu(f) \in \boldsymbol{M}$. Condition (a) will guarantee that at least one mean exists, but generally there will be many. We therefore introduce

Definition 4. $M_\mu(f)$ is that subset of $\boldsymbol{M}$ on which $V_\mu(f,\cdot)$ takes on its minimum value. (*Set of means of f.*)

Definition 5. If $M_0, M_1, \ldots$ are subsets of $\boldsymbol{M}$, we say that $M_n \to M_0$, $n = 1,2,\ldots$, if every open set containing M_0 contains almost all the M_n.

The connection between means and variances is given by the following lemma.

Lemma 1. If $V_{\nu_n}(f,t) \to V_\nu(f,t)$ uniformly in t, then $M_{\nu_n}(f) \to M_\nu(f)$.

Proof. Since f is fixed, it will be omitted from the formulas.

Let $M_\nu \subseteqq O$, O open. Then, by (a), V_ν takes on a minimum value v on $\tilde{O}$. Let w be the minimum of V_ν on all of $\boldsymbol{M}$. Since no mean is in $\tilde{O}$, $v > w$.

Let $m \in M_\nu$.

$$V_{\nu_n}(t) - V_{\nu_n}(m) = [V_{\nu_n}(t) - V_\nu(t)] + [V_\nu(t) - V_\nu(m)] + [V_\nu(m) - V_{\nu_n}(m)].$$

Choose n large enough so that $|V_{\nu_n}(t) - V_\nu(t)| < (v - w)/2$, for all $t \in \boldsymbol{M}$. This is possible by the uniform convergence. If $t \in \tilde{O}$, then the second term is at least $v - w$, and hence the right side is positive. Hence $M_{\nu_n} \cap \tilde{O} = \emptyset$ for sufficiently large n.

Lemma 2. If $\nu_n \to \nu$ (in the sense of Def. 2), then $V_{\nu_n}(f,t) \to V_\nu(f,t)$ uniformly in t.

This is an immediate consequence of (c).

Our program is now as follows: We introduce the measures $\mu_{n,x}$ as before. We show that they converge to measures μ_x a.e. (in the sense of Definition 2). Then we will know that $M_{\mu_{n,x}}(f) \to M_{\mu_x}(f)$ a.e. We will then have an abstract version of (I′), where $M_{\mu_{n,x}}(f)$ is the abstract analogue of $m_{\mu_{n,x}}(f)$ and $M_{\mu_x}(f)$ takes the place of $a(x)$,

Lemma 3. There exist measures μ_x, depending on x in general, such that $\mu_{n,x} \to \mu_x$ a.e. And if T is ergodic, $\mu_x = \mu$ a.e.

Proof. Let f_i be the characteristic function of $S_i \in A$.

$$\mu_{n,x}(S_i) = (1/n)[f_i(x) + f_i(Tx) + \ldots + f_i(T^{n-1}x)].$$

Hence $\mu_{n,x}(S_i) \to \nu_x(S_i)$ a.e., by the classical theorem. Since the set of points where convergence fails for a given S_i has measure 0, the union of these (denumerably many) "failure" sets also has measure 0. Hence its complement has measure 1. Thus $\mu_{n,x}(S_i) \to \nu_x(S_i)$ *for all i* a.e. Clearly, $\nu_x(\emptyset) = 0$, $\nu_x(\boldsymbol{X}) = 1$, $\nu_x(S_i) \geqslant 0$. We can also prove that ν is countably additive a.e., using Theorem 2.1, p. 465 in [1] together with properties CE_3 and CE_5 (p. 23). Hence ν_x is a measure on $\mathbf{A}$ a.e.

Due to (b), ν_x can be extended to a measure μ_x on $\mathbf{S}$, whenever ν_x is a measure on $\mathbf{A}$. Otherwise we arbitrarily let $\mu_x = \mu$. Then $\mu_{n,x} \rightarrow \mu_x$ (in the sense of definition 2) a.e.

If T is ergodic, then $\nu_x(S_i) = \mu(S_i)$ a.e., and since the extension to S is unique, $\mu_x = \mu$ a.e.

ERGODIC THEOREM.

(I*) $M_{\mu_{n,x}}(f) \rightarrow M_{\mu_x}(f)$ a.e.

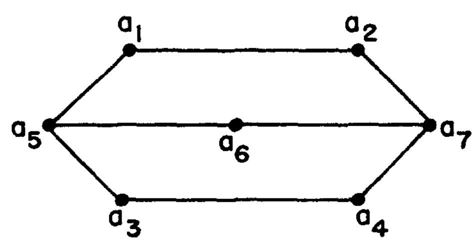
(II*) $M_{\mu_{Tx}} = M_{\mu_x}$ a.e.

(IV*) If T is ergodic, $M_{\mu_x}(f) = M_\mu(f)$ a.e.

PROOF. (I*) and (IV*) are immediate consequences of Lemmas 1, 2, 3. (II*) follows from the fact that by the classical theorem $\nu_{Tx}(S_i) = \nu_x(S_i)$ a.e. The three parts of the theorem are abstract analogues of the corresponding classical results.

The abstract analogue of (III') is not immediately clear. While $m_\mu(f)$ surely goes over into $M_\mu(f)$, what happens to $m_\mu(a)$? We have represented $a(x)$ by $M_{\mu_x}(f)$, and we must somehow take a mean of a set of sets of means.

We will show in a simple example that serious difficulties arise. Let $\boldsymbol{X} = \boldsymbol{R} = \boldsymbol{M}$ be the 7-point space shown in the figure. The measure μ



assigns weight 1/4 to points a_1,a_2,a_3,a_4.

Let f be the identity function. We define a metric δ by taking the length of each segment in the figure to be 1. We introduce the natural variance:

$$V_\mu(f,t) = \int \delta^2(f(x),t)\, d\mu(x) = \tfrac{1}{4} \sum_{i=1}^{4} \delta^2(a_i,t).$$

Then $M_\mu(f) = \{a_5,a_7\}$.

First let us take T to be the cyclic permutation (1 2 3 4). This is measure preserving and ergodic. Hence $\mu_x = \mu$ a.e. So we must find a mean of four sets (for $x = a_1, a_2, a_3, a_4$) each of which is $\{a_5, a_7\}$. Presumably we must make a selection from each set, and then take the mean of the selected set. But if a_5, a_7 are each selected twice (hence have weight 1/2 assigned), then the mean of the selected set is a_6, which is not in $M_\mu(f)$. We might think in this case that "the right selection" would work. But even this need not be the case.

Let T be the permutation (1 3) (2 4). This is measure preserving, but not ergodic. If $x = a_1$ or a_3, μ_x will assign 1/2 each to a_1 and a_3. Hence $M_{\mu_x} = \{a_5\}$. If $x = a_2$ or a_4, $M_{\mu_x} = \{a_7\}$. Since each turns up with probability 1/2, the mean of the M_{μ_x} sets *must* be a_6 in this case. But $a_6 \notin M_\mu(f)$.

Another version of (III) may also be considered. We note that in the ergodic case (III) is a simple consequence of (IV). But the result, before the obvious simplification, appears as

$$\int a \, d\mu = \int\int f \, d\mu \, d\mu. \tag{III''}$$

This would suggest that each mean of the sets $M_{\mu_x}(f)$ should be a mean of the set M_μ. This would surely be true in the ergodic case, where $\mu_x = \mu$ a.e. But that even this will not always be true is seen in another example. In the above figure, let μ assign weight 1/6 to the points a_1, a_3, and a_5, and weight 1/8 to the others. Let T be the permutation (1 3 5) (2 4 6 7). Then for three of the points, having total weight 1/2, μ_x assigns 1/3 each to a_1, a_3, a_5, and hence $M_{\mu_x}(f) = \{a_5\}$. For the other points, similarly, $M_{\mu_x}(f) = \{a_7\}$. Hence the mean of these sets will be a_6. But $M_\mu(f) = \{a_5\}$.

It is remarkable that analogues of (I), (II), (IV) were proved without making any assumption on how $V_\mu(f, \cdot)$ depends on f. We couldn't possibly hope to prove an analogue of (III) without this information, since in (III) we compare means of two different functions with respect to the same measure. But the above examples show that we may also have to place severe restrictions on the space M. This problem should be investigated.

Another avenue of promising research is the attempt to weaken the requirement for the existence of variances, by a method of truncation (for example, see [2]).

Let us now examine a few applications of the abstract ergodic theorem. Several examples for the case where $\boldsymbol{R}$ is denumerable and T ergodic were discussed in ref. [4]. Let us therefore concentrate on a non-denumerable $\boldsymbol{R}$.

For the first three examples we will choose as our measure space the interval [0,1] with a regular Borel measure. T is any transformation preserving this measure. We will start with a classical example, and go on to less classical ones.

Example 1. f is a continuous real-valued function, and $\boldsymbol{M} = \boldsymbol{R}$ is an interval. The variance is $V_\mu(f,t) = \int[f(x) - t]^2 \, d\mu(x)$. To satisfy (b) we choose the algebra generated by the rational intervals, and then (a) and (c) will be satisfied. This example is, of course, coverd by the classical theorem. What does our ergodic theorem say about it?

We note $M_\nu(f)$, for any measure ν, consists of the single point $\int f \, d\nu$. For unit sets convergence according to Def. 5 coincides with ordinary convergence. Hence we obtain (I), (II), and (IV). We also obtain the following representation for the limiting function:

$$a(x) = \int f(y) \, d\mu_x(y).$$

This result is not usually included in the ergodic theorem — indeed, it fails to hold in certain pathological situations. That the representation exists for our example may be seen from Theorem 9.5, p. 31 in [1]. Our abstract ergodic theorem provides an independent proof of this fact.

Example 2. A semi-classical example may be obtained by keeping f, but choosing a different variance. We have considerable freedom in our choice. For example, $V_\mu(f,t) = \int|f(x) - t| \, d\mu(x)$ will produce medians as means. Thus we have an ergodic theorem for medians. It should be noted that medians need not be unique, and hence even in this semi-classical example convergence must be taken in the sense of Def. 5.

Example 3. We choose a continuous function f whose range lies in a compact metric space $\boldsymbol{M}$. If δ is the metric of the space, $V_\mu(f,t) = \int\delta^2(f(x),t) \, d\mu(x)$ is a natural choice for the variance. All our conditions are satisfied, and hence our ergodic theorem applies to the metric-valued function.

Example 4. The choice of the real measure space for Example 3 was inessential. We could, for example, have chosen a locally compact space which can be denumerably generated. The only difference is that to satisfy (B) we must choose the algebra generated by a set of generators of the measurable sets together with the sets $\{x|\delta^2(f(x),t) \leqslant r\}$ for all rational r. Then all our conditions are satisfied, and we have a completely abstract example.

References

1. Doob, J. L., "Stochastic Processes." Wiley, New York, 1953.
2. Fréchet, M., *Les Éléments Aléatoires de nature quelconque dans un espace distancié*, Institut Henri Poincaré, vol. X, no. IV, 1948.
3. Halmos, P. R., "Lectures on Erdogic Theory." The Mathematical Society of Japan, Tokyo, 1956.
4. Kemeny, J. G., Generalized random variables. *Pacific J. Math.* **9**, No. 4, 1179–1189 (1959).